

doi: 10.7690/bgzdh.2019.12.013

# 基于 SVM 决策树的数据链识别分类

杨 咪, 王安丽, 胡 正

(西北工业大学航空学院, 西安 710072)

**摘要:**为解决数据链信号识别分类的问题,提出一种基于支持向量机(support vector machine, SVM)决策树的数据链识别分类方法。通过分析美军常用的数据链通信信号特征,采用小波变换法分析数据链的特征信息,得出小波系数与信号能量分布的关系,根据 SVM 算法原理,构建目标特征模型,对信号特征量进行识别分类,对 SVM 分类器的关键参数进行优化设计,并与 BP 神经网络算法进行对比实验仿真。结果表明:SVM 决策树网络分类器在进行收敛速度和准确率表现优异,能改善分类识别效能。

**关键词:**数据链;小波变换;支持向量机;决策树;模式识别

**中图分类号:**TP391 **文献标志码:**A

## Classification and Recognition of Data Link Based on SVM Decision Tree

Yang Mi, Wang Anli, Hu Zheng

(School of Aeronautics, Northwestern Polytechnical University, Xi'an 710072, China)

**Abstract:** In order to solve the problem of data link signal identification and classification, a data link recognition and classification method based on Support Vector Machine (SVM) decision tree is proposed. By analyzing the characteristics of the data link communication signals commonly used by the US army, the wavelet transform method is used to analyze the characteristic information of the data link, and the relationship between wavelet coefficients and signal energy distribution is obtained. According to the principle of SVM algorithm, the target feature model is constructed to identify and classify the signal feature quantities. The key parameters of the SVM classifier are optimized and compared with the BP neural network algorithm. The results show that the SVM decision tree network classifier performs well in convergence speed and accuracy, and can improve classification and recognition performance.

**Keywords:** data link; wavelet transform; SVM; decision tree; pattern recognition

## 0 引言

先进的数据链为不同作战体系之间进行信息无缝交互、高度互操作提供了物理基础。数据链信号的识别分类是信号检测、盲接收、电子对抗等作战行动中一个十分重要的研究内容<sup>[1-2]</sup>。数据链路信号通过天线、雷达、侦查卫星以及各种探测传感器获得,通信信号在传播过程中,由于区域的开放性和信息的隐蔽性,导致数据链信号的识别分类面临挑战。一方面由于信息化战争的复杂性,无法获取到足够多的数据信号样本,必须在小样本训练环境下得到分类模型;另一方面对数据链信号的分类必须要在短时间内完成,以满足战争实时性的要求。

对于数据链的识别分类研究,文献[3]采用模糊评价理论对 link11 数据链进行了参数识别。该方法在选取不同模糊评价矩阵的算子和指标参数时,对评价结果影响很大,易产生误判。文献[4]通过对 link16 数据链信号的 3 个特征量进行识别进而确定数据链信息,但对数据链信号缺乏系统的分析。文

献[5]采用支持向量机算法对数据链通信信号的调制方式进行了识别,但没有完整地分析数据链信号整体特征。

笔者针对数据链信号的识别分类问题,提出采用多频滤波、小波变换等对数据链信号进行分析处理的方法,获取信号的频率与能量的小波特征向量。在传统决策树分类模型的基础上,采用 SVM 决策树算法克服决策树本身对门限敏感与错误传递的缺陷,构建系统的 SVM 决策树数据链网络分析模型,对提取的特征向量进行训练,从而对复杂环境中的多种数据链信号快速识别分类。

## 1 军用数据链信号的特征分析与提取

美军使用的数据链信号的调制方式以及采用的频率带等相关信息如表 1 所示。

对数据链信号的分析处理包括时域分析、频域分析、功率谱分析以及时频率分析等,不同的分析方法可得到不同的特征信息,但对于这些特征信息

收稿日期:2019-08-23;修回日期:2019-09-28

基金项目:国家自然科学基金(61573283)

作者简介:杨咪(1995—),女,陕西人,博士,从事多飞行器协同控制、信息融合、任务分配等研究。E-mail: yangmi@mail.nwpu.edu.cn.com。

的使用尚处于分析阶段，在辅助决策等领域应用较少。而小波变换是一种变时频分析方法，通过选择合适的变换能充分突出问题某些方面特征，是进行信号时频分析和处理的有利工具，因此，笔者采用小波变换法分析数据链的特征信息。

表 1 美军常用数据链信息汇总

数据链类型	调制方式	频带	访问协议
Link11	$\pi/4$ -DQPSK	2~30 MHz 与 225~400 MHz 969~1 008 MHz、	TDMA
Link16	MSK	1 053~1 065 MHz 以及 1 113~1 206 MHz	TDMA
Link22	QPSK、8PSK	3~30 MHz 与 225~400 MHz	TDMA 和动态 TDMA
CDL	BPSK、OQPSK	X 波段和 ku 波段	TDMA

### 1.1 数据链信号的频率分段与分解

将接收到的数据链信号经多频带滤波器进行频率带划分，由表 1 数据链可划分为 7 个连续的波段：

- 第 1 段：2~30 MHz；
- 第 2 段：225~400 MHz；
- 第 3 段：969~1 008 MHz；
- 第 4 段：1 053~1 065 MHz；
- 第 5 段：1 113~1 206 MHz；
- 第 6 段：X 波段；
- 第 7 段：ku 波段。

对每段频率类的信号  $x_i(t)$  进行连续的离散小波分解(DWT)：

- 第 1 层：第一次分解后得到小波系数  $a_1$  与  $c_1$ ；
- 第 2 层：对  $a_1$  进行分解得到  $a_2$  与  $c_2$ ；
- 第  $i$  层：对上一层的  $a_{i-1}$  进行分解得到下一层的  $a_i$  与  $c_i$ 。

则每个波段的信号可以表示为不重叠的各个子带的小波之和，表示为  $[a_n, c_n, \dots, c_{i-1}, \dots, c_1]$  ( $n$  为分层数)。

### 1.2 构建小波系数与各频率的能量对应关系

在时域上，假设各频率段信号输入为  $x_i(t)$ ，其在  $L^2(\mathbb{R})$  空间定义范数的平方表示频段内信号的能量为：

$$\|X\|_2^2 = \int_{\mathbb{R}} |x_i(t)|^2 dt. \quad (1)$$

对于选取的小波函数有：

$$\|X\|_2^2 = \sum_{k \in \mathbb{Z}} |\langle x, \varphi_{j,k} \rangle|^2. \quad (2)$$

式中： $j$  为窗口系数； $k$  为平移值。

由式(1)、(2)可建立小波系数与时域能量之间的关系。

设小波系数为  $V_{j,k}(a_{j,k}$  或  $c_{j,k})$ ，对应的能量为  $E_{j,k}$ ，有

$$E_{j,k} = \sum_{i=1}^N |s_{j,k}|^2, j, k \in \mathbb{Z}^+. \quad (3)$$

式中： $N$  为抽样点数； $j$  为窗口参数； $k$  为平移值。

### 1.3 提取各个频率段的信号特征向量

提取各个频率段的信号特征向量的步骤如下：

1) 使用雷达、天线、卫星、传感器等检测监控设备获取原始信号；

2) 将原始信号通过多频带滤波器，得到多段连续谱信号  $x_i(t)$ ；

3) 对  $x_i(t)$  进行抽样，获取离散的各频率段信号序列  $X_i(n)$ ；

4) 对离散信号序列  $X_i(n)$  进行多层小波分解，得到 1 维小波系数，并由式(3)计算小波系数与频率点对应的能量关系；

5) 构造各个频率段信号的特征向量

$$P_i = [E_{c1}, E_{c2}, \dots, E_{cn}]. \quad (4)$$

式中  $n$  为小波分解的层数。

## 2 数据链识别的 SVM 决策树网络算法

### 2.1 SVM 算法原理

SVM 的基本模型定义为特征空间上间隔最大的线性分类器，学习策略是间隔(如图 1 中的 Gap)的最大化，在小样本分析时优势明显。同时，针对低维度空间线性不可分类的情形，选择合适的核函数 Kernel 方程，隐式地将非线性的训练数据映射到高维空间(如图 2)，而不增加可调参数的个数，在高维空间进行识别分类。

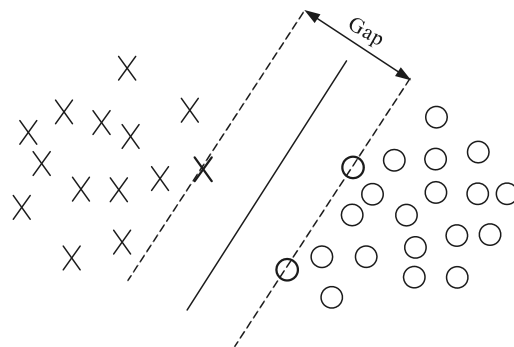


图 1 间隔最大线性分类器

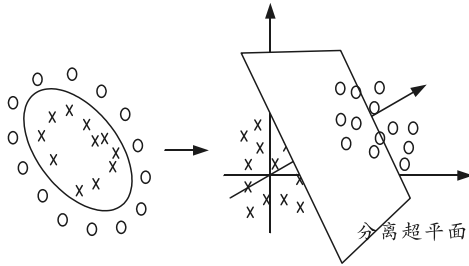


图 2 应用核函数的高维映射

SVM 算法满足结构风险的最优化原则如式(5), 分类使用的超平面可使几何间隔最大如式(6):

$$\min_{f \in F} \frac{1}{N} \sum_{i=1}^N L(y_i, f(x_i)) + \lambda J(f). \quad (5)$$

式中:  $x_i$  为输入;  $y_i$  为对应输出;  $f$  为选择模型;  $F$  为假设空间。

$$\tilde{\gamma} = \gamma \gamma = \frac{\hat{\gamma}}{\|w\|}. \quad (6)$$

式中:  $\tilde{\gamma}$  为几何间隔;  $|\gamma|=1$ ;  $\hat{\gamma}$  为函数间隔;  $w$  为参数矩阵。

### 2.2 SVM 分类器的设计

假定信号采集的  $n$  组  $C$  维的特征值样本表示为  $(x_i, y_i), i=[1, n]$ ;  $x \in R_c, y=\{1, -1\}$ , 上述 SVM 分类器的目标函数可表示为:

$$\min \frac{1}{2} \|w\|^2 \quad \text{s.t.}, y_i(w^T x_i + b) \geq 1, i=1, \dots, n. \quad (7)$$

实际情况下, 由于战场环境的开放性与复杂性, 获得的信号数据中会有噪声、干扰波等。为了优化干扰源的影响, 引入松弛变量与惩罚因子, 同时简化计算难度, 通过拉格朗日对偶性变换到对偶变量的优化问题, 将式(7)进一步优化得到目标方程组如下式所示:

$$\begin{aligned} \min & \frac{1}{2} (w) + C \sum_{i=1}^n \xi_i, \\ \text{s.t.}, & y_i(w^T x_i + b) \geq 1 - \xi_i, i=1, \dots, n, \\ & \xi_i \geq 0, i=1, \dots, n. \end{aligned} \quad (8)$$

其中:  $w, b$  为待求系数;  $\xi_i$  为松弛变量, 表示因为各种原因不能被正确分类的样本集合;  $\xi$  的取值可由经验值得到;  $C$  为对应的惩罚参数, 控制松弛变量的影响。

由于得到的特征向量为多维空间向量, 在低维空间线性不可分, 因此, 引入高斯核函数对目标函数的求解进行优化, 其数学表达式如下:

$$K(x_1, x_2) = \exp\left(-\frac{\|x_1 - x_2\|^2}{2\sigma^2}\right). \quad (9)$$

其中:  $\|x_1 - x_2\|$  为向量间的距离;  $\sigma$  为给定常数, 可采用优化算法进行优化选取。

军用数据链系列发展至今已经有数 10 种之多, 属于多类分类问题, 在进行 SVM 分类器设计时可以进行 SVM 分类器的级联设计或网络化设计, 每次只进行一次二类分类, 从而可以将多类分类问题转化为多步的二类分类问题来完成。

采用 SVM 决策树理论, 对式(4)的特征向量进行训练, 经过逐级的判决与分类后, 可得到分类结果。具体流程如图 3 所示。

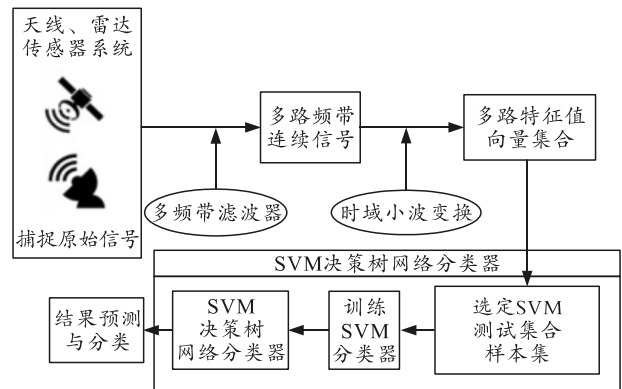


图 3 SVM 决策树网络数据链识别分类系统

### 2.3 SVM 分类器参数优化

SVM 分类器中惩罚系数  $C$  与高斯核函数中  $\sigma$  对于 SVM 分类器的最后判断结果影响重大。  $C$  用于控制目标函数式(8)中寻找分类的最大超平面与保证数据点偏差量最小之间的权重系数;  $\sigma$  控制映射的比例, 当  $\sigma$  过大, 高次特征上的权重衰减得过快, 如果选得过小, 则会将任意的数据映射为线性可分<sup>[7]</sup>, 不能达到识别分类的目的。这些参数通常需要提前给定, 受经验影响很大, 往往不能得到最优的参数结果; 因此, 需要采用合适的算法对参数进行优化设计。

对于 SVM 参数的设计, 文献[5]采用 PSO 算法对  $C$  和  $\sigma$  进行优化设计, 能够快速收敛, 但容易陷入局部最优, 无法得到全局最优解; 文献[8]采用遗传算法对参数进行优化, 同样容易陷入局部最优解, 且训练时间长, 不满足实时性要求; 文献[9]在传统的启发式算法的基础上, 提出了一种采用改进的网格搜索法, 先在较大的范围进行粗搜索, 确定最优解的区间, 然后采用小步距进行精确收缩, 能在较短时间内达到最优解, 相较于 GA 算法与 PSO 算法所用时间更短, 精确度更高。笔者采用改进的网格搜索法对上述的参数  $C$  和  $\sigma$  进行优化, 其流程如图 4 所示。

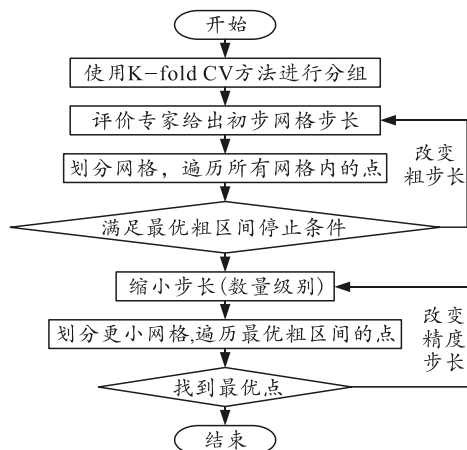


图 4 基于改进的网格搜索法的极值寻优流程

### 3 实验仿真与分析

#### 3.1 实验仿真设定

使用多台飞利浦 PM5518-TX 电视信号发生器，在频率为 32~1 300 MHz 内，设定指定强度发送连续波段频率模拟数据链信号，采用多极微波天线接收器接收不同频段的通信信号，接收的信号选取如下 5 段进行小波变换与数据处理<sup>[10-11]</sup>。第 1 段，2~30 MHz；第 2 段，225~400 MHz；第 3 段，969~1 008 MHz；第 4 段，1 053~1 065 MHz；第 5 段，1 113~1 206 MHz。得到的特征向量集合后，分别经过 SVM 决策树网络与 BP 神经网络进行识别分类，统计两者在不同信噪比下的收敛速度和平均准确率。

#### 3.2 仿真分析

笔者将 SVM 决策树网络仿真结果和传统的模式识别算法中的 BP 神经网络算法进行对比，得出统计特性。图 5 是在不同的样本容量下 2 种算法收敛时间的对比情况，可以看出：应用 SVM 决策树设计的分类算法所用的收敛时间较短，且 BP 神经网络在样本容量较大时，收敛时间急剧增加，而 SVM 决策树分类器随着样本容量的增加，收敛时间增长不剧烈，且在小容量样本中能做到快速收敛。

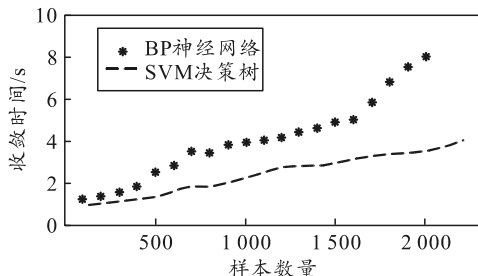


图 5 2 种分类算法的收敛时间对比

假定在识别过程中各个信号出现的概率均等，

统计在不同信噪比下，2 种算法对信号的平均正确识别率如图 6 所示。在信噪比过低时，2 种算法正确识别率都较低，在超过一定信噪比时，SVM 决策树算法明显优于 BP 神经网络算法，在高信噪比时，2 种算法都能准确分辨数据链信号，但基于 SVM 决策树算法总体优于 BP 神经网络算法。

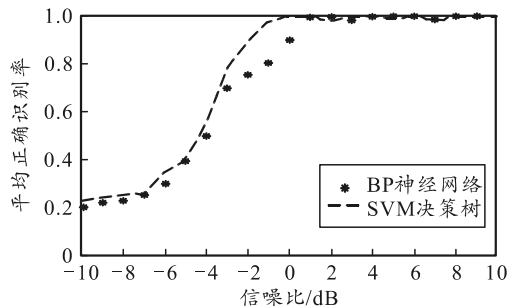


图 6 2 种分类算法的正确识别率对比

讨论不同  $C$  与  $\sigma$  对 SVM 决策树算法的影响时，随机选择 6 组不同  $(C, \sigma)$  对，在信噪比为 -1.2 dB 的条件下，统计平均正确识别率如表 2 所示，在惩罚因子  $C=2$ ，高斯核参数  $\sigma=1$  时效果较好。

表 2 不同  $(C, \sigma)$  SVM 分类器平均正确率 %

$(C, \sigma)$ 取值	平均正确率
(2,2)	81.120 3
(2,1)	88.231 0
(5,2)	65.330 8
(1,8)	50.125 0
(3,9)	45.239 8
(4,4)	65.102 0

采用基于改进的网格搜索法进行进一步优化，经过 580 次迭代后收敛，在  $(C, \sigma)=(2.15, 1.24)$  时取得最优解，平均正确率达到 94.273 5%，满足准确率要求。

### 4 结束语

笔者通过分析美军常用的数据链通信信号特征，包括调制方式、通信频段、访问协议等，采用小波变换方式对捕捉到的原始信号进行特征分析，得出小波系数与信号能量分布的关系。同时采用基于 SVM 决策树算法对信号特征量进行识别分类，并对 SVM 分类器的关键参数进行优化设计，改善了分类识别效能。仿真分析结果显示，SVM 决策树分类器在收敛速度、识别分类准确率方面满足要求。

### 参考文献：

[1] POLYDOROS A, KIM K. On the detection and classification of quadrature digital modulations in broad-band noise[J]. IEEE Transactions on Communications, 1990, 38(8): 1199-1211.

## 5 结束语

笔者研究了主被动一体化隔振平台, 针对目前通用的力控制模型, 将位移跟踪模型与特征建模控制方法相结合, 分析设计了基于位移跟踪控制的模型, 并进行仿真验证。数学仿真结果表明: 该控制器对于干扰力有明显的抑制作用, 并能有效地减小上层平台位移, 证明了黄金分割控制以及逻辑微分控制的优越性, 为进一步研究载荷隔振平台在实际航空飞行器中的应用打下了基础。

## 参考文献:

- [1] 张磊, 付永领, 刘永光, 等. 主动隔振技术及其应用与发展[J]. 机床与液压, 2005(2): 9-12.
  - [2] KARNOPP D. Active and Semi-Active Vibration Isolation[J]. Current Advances in Mechanical Design Production VI, 1995, 117(B): 409-423.
  - [3] 郑钢铁. 空间飞行器隔振与振动抑制研究进展[C]//全国振动理论及应用学术会议, 2007.
  - [4] 卢兵仔, 罗均, 唐文, 等. 超小型无人旋翼飞行器减振增稳云台的设计[J]. 机械制造, 2007, 45(7): 24-26.
  - [5] 肖斌. 柴油机双层隔振台架主动隔振技术研究[D]. 哈尔滨: 哈尔滨工程大学, 2008.
  - [6] KIM S M, ELLIOTT S J, BRENNAN M J. Decentralized control for multichannel active vibration isolation[J]. IEEE Transactions on Control Systems Technology, 2001, 9(1): 93-100.
  - [7] 吴宏鑫, 沈少萍. PID 控制的应用与理论依据[J]. 控制工程, 2003, 10(1): 39-44.
  - [8] WANG L J, MENG B. Characteristic model-based control of robotic manipulators with dynamic uncertainties[J]. Sci China Inf Sci, 2017, 60(7): 079201.
  - [9] 吴宏鑫, 解永春, 李智斌, 等. 基于对象特征模型描述的智能控制[J]. 自动化学报, 1999, 25(1): 9-17.
  - [10] QIU Z, ZHANG X, WU H, et al. Optimal placement and active vibration control for piezoelectric smart flexible cantilever plate[J]. Journal of Sound Vibration, 2007, 301(3): 521-543.
  - [11] 吴宏鑫, 王迎春, 邢琰. 基于智能特征模型的智能控制及应用[J]. 中国科学(技术科学), 2002, 32(6): 805-816.
  - [12] 孟斌, 吴宏鑫. 黄金分割控制的收敛性和稳定性研究[J]. 宇航学报, 2009, 30(5): 2128-2132.
  - [13] 史敬灼, 尤冬梅. 超声波电机黄金分割自适应转速控制[J]. 电工技术学报, 2013, 28(6): 59-65.
  - [14] LI H, LIU H, GAO H, et al. Reliable Fuzzy Control for Active Suspension Systems With Actuator Delay and Fault[J]. IEEE Transactions on Fuzzy Systems, 2012, 20(2): 342-357.
  - [15] 吴宏鑫. 全系数自适应控制理论及其应用[M]. 北京: 国防工业出版社, 1990: 121-146.
  - [16] 王晓磊, 吴宏鑫. 挠性航天器振动抑制的自适应方法及实验研究[J]. 宇航学报, 2005, 26(3): 275-281.
  - [17] WU H, HU J, XIE Y. Characteristic Model-Based All-Coefficient Adaptive Control Method and Its Applications[J]. IEEE Transactions on Systems Man & Cybernetics Part C, 2007, 37(2): 213-221.
  - [18] 张旭辉, 刘永光, 付永领. 磁致伸缩作动器结构优化设计[J]. 压电与声光, 2009, 31(3): 377-380.
  - [19] 张磊, 付永领, 刘永光, 等. 磁致伸缩作动器的建模与控制研究[J]. 压电与声光, 2005, 27(5): 42-45.
- \*\*\*\*\*
- (上接第 57 页)
- [2] DOBRE O A, ABDI A, BAR-NESS Y, et al. Survey of auto-matic modulation classification techniques: classical approaches and new yrends[J]. IET Communications, 2007, 1(2): 137-156.
  - [3] 范玲. Link-11 数据链信号的识别方法研究[D]. 西安: 西安电子科技大学, 2014.
  - [4] 李国鑫. Link16 数据链信号的识别算法研究[D]. 西安: 西安电子科技大学, 2014.
  - [5] 甘新泰, 程伟, 杨瑞娟, 等. 基于支持向量机分类器的数据链信号调制方式识别[J]. 空军预警学院学报, 2015, 29(1): 38-42.
  - [6] 刘国栋, 许静. 基于 SVM 方法的神经网络呼吸音识别算法[J]. 通信学报, 2014, 35(10): 218-222.
  - [7] Choosing Multiple Parameters for Support Vector Machines[J]. Olivier Chapelle, Vladimir Vapnik, Olivier Bousquet, Sayan Mukherjee. Machine Learning, 2002, 46(1): 131-159.
  - [8] 连可, 陈世杰, 周建明, 等. 基于遗传算法的 SVM 多分类决策树优化算法研究[J]. 控制与决策, 2009, 24(1): 7-12.
  - [9] 王健峰, 张磊, 陈国兴, 等. 基于改进的网格搜索法的 SVM 参数优化[J]. 应用科技, 2012, 39(3): 28-31.
  - [10] 刘红军. 美军公用数据链发展概述[J]. 中国电子科学研究院学报, 2008, 3(6): 608-612.
  - [11] 周子栋, 陈自力, 高喜俊, 等. 无人机数据链常面临的干扰类型的自动识别[J]. 计算机测量与控制, 2015, 23(11): 3780-3782.